# This Page Is Inserted by IFW Operations and is not a part of the Official Record

# **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

## IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.

#### PATENT ABSTRACTS OF JAPAN

(11) Publication number: 10254496 A

(43) Date of publication of application: 25 . 09 . 98

(51) Int. CI

G10L 5/06 G06F 15/18 G10L 3/00

(21) Application number: 09056162

(22) Date of filing: 11 . 03 . 97

(71) Applicant:

MITSUBISHI ELECTRIC CORP

(72) Inventor:

**ABE YOSHIHARU** 

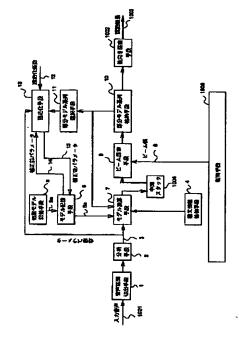
#### (54) SPEECH RECOGNITION METHOD

#### (57) Abstract:

PROBLEM TO BE SOLVED: To make it possible to reduce a search volume and also improve accuracy of speech recognition by determining a likelihood of an input speech to model group composed of adapted model groups by a search processing and configuring a system to recognize the speech based on this likelihood.

SOLUTION: A model calculating means 7 receives and input for one frame, and applies a parameter of a model in accordance with syntax information for each frame to the input and calculates the likelihood of the inputted feature parameter 3. Next, a partial model group selection means 11 selects a partial model group as information used for an adaptation at an adapting means 13. Next, after adaptation has been completed for all the selected models, the adapting means 14 adopts the parameters obtained from the result of adaptation as corrected parameters and replaces the parameters of a storage means 6 with the corrected parameters 15. And backward searching is operated by a backward searching means 1002 and a recognition result 1003 is obtained.

COPYRIGHT: (C)1998,JPO



#### (19)日本国特許庁(JP)

## (12) 公開特許公報(A)

(11)特許出願公開番号

### 特開平10-254496

(43)公開日 平成10年(1998) 9月25日

(51) Int.Cl. <sup>6</sup>	識別記号	F I	
G10L 5/06		G 1 0 L 5/06 D	
G06F 15/18	560	G 0 6 F 15/18 5 6 0 G	
G 1 0 L 3/00	5 3 1	G 1 0 L 3/00 5 3 1 D	
		5 3 1 K	

審査請求 未請求 請求項の数11 〇1、(全 15 頁)

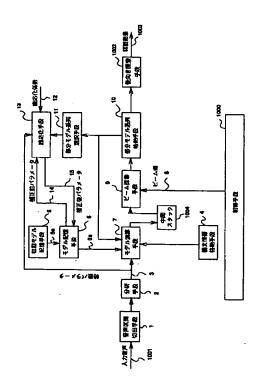
		<b>一种红朗</b> 水	术明水 明水填の数II UL (全 15 貝)		
(21)出願番号	特顯平9-56162	(71)出顧人	000006013 三菱電機株式会社		
(22)出願日	平成9年(1997)3月11日	(72)発明者	東京都千代田区丸の内二丁目2番3号 阿部 芳春 東京都千代田区丸の内二丁目2番3号 三 菱電機株式会社内		
		, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,			
		(74)代理人	弁理士 宮田 金雄 (外2名)		

#### (54) 【発明の名称】 音声認識方式

#### (57) 【要約】

【課題】 探索を用いる音声認識において、入力音声の途中でビーム探索の閾値を変更する従来のビーム探索は、探索量を削減することができるが、認識に用いるモデルのパラメータは一定であり、このようなパラメータが一定のモデルで認識を行うため認識精度の向上は得られない。この発明が解決しようとする課題は、探索量を削減することができ、かつ認識の精度を向上させることである。

【解決手段】 入力音声の各フレームで入力音声にモデルを適応化し、この適応化されたモデルに対する入力音声の尤度を探索処理により求め、この尤度に基づき音声認識を行う。



【特許請求の範囲】

【請求項1】 入力音声の各フレームで入力音声にモデルを適応化し、この適応化されたモデルの系列からなるモデル系列に対する入力音声の尤度を探索処理により求め、この尤度に基づき音声認識を行うことを特徴とする音声認識方式。

【請求項2】 上記探索処理としてビーム探索を用いることを特徴とする請求項1記載の音声認識方式。

【請求項3】 入力音声の途中のビーム内の仮説の尤度 探索の閾値の制御関数に の大きい方からN個の仮説を用いて、上記モデルを入力 10 回帰分析を用いている。 音声へ適応化することを特徴とする請求項2記載の音声 認識方式。 発表会 講演論文集 1

【請求項4】 入力音声の途中のビーム内の仮説の尤度の大きい方からN個の仮説を用いるとともに、尤度に応じた重みによる加重を用いて上記モデルを入力音声へ適応化することを特徴とする請求項2記載の音声認識方式。

【請求項5】 上記モデルを音韻のモデルとしたことを 特徴とする請求項1ないし4いずれかに記載の音声認識 方式。

【請求項6】 上記モデルを音韻境界のモデルとしたことを特徴とする請求項1ないし4いずれかに記載の音声認識方式。

【請求項7】 上記モデルを音韻のモデルおよび音韻境 界のモデルとしたことを特徴とする請求項1ないし4い ずれかに記載の音声認識方式。

【請求項8】 上記音韻のモデルをセミ連続分布のモデルとし、音韻のモデルの分岐係数だけを適応化することを特徴とする請求項5または7いずれかに記載の音声認識方式。

【請求項9】 上記音韻境界のモデルをセミ連続分布のモデルとし、音韻境界のモデルの分岐係数だけを適応化することを特徴とする請求項6または7いずれかに記載の音声認識方式。

【請求項10】 上記音韻境界のモデルの入力音声への 適応化において、入力音声の途中のビーム内の仮説の尤 度の大きい方からN個の仮説を選択する際に、モデル間 の遷移を有する部分モデル系列の仮説を選択することを 特徴とする請求項6記載の音声認識方式。

【請求項11】 ビーム探索の探索の幅をフレームごと に適応化係数を用いて変化させることを特徴とする請求 項2ないし4いずれかに記載の音声認識方式。

【発明の詳細な説明】

[0001]

【発明の属する技術分野】この発明は、音声認識の精度 改善と、探索量の削減に関する。

[0002]

【従来の技術】従来、音声認識の探索量の削減の手法として、ビーム探索が行われている。日本音響学会 平成元年度春季研究発表会 講演論文集 I (平成元年3

月)、5~6頁「DPビームサーチの関値を入力音声の途中で変更する方法の検討」には、フレーム同期型のDPマッチングにおいて、ビーム探索の関値を入力音声の途中で変化させることで、探索量を減少させる方法が提案されている。また、特開平6-282295号公報には、観測可能な特徴量を入力とする制御関数を用いてビーム探索の探索範囲を適応的に変化させることで、探索量を減少させる方法が開示されている。ここで、ビーム探索の関値の制御関数には、ニューラルネットおよび重回帰分析を用いている。

【0003】一方、日本音響学会 平成8年度秋季研究 発表会 講演論文集 I (平成8年9月)、117~118頁「音声認識のためのN-bestに基づく話者適応 化」には、教師なし話者適応化(即時適応化)の方法として、Nベストビタビ認識の結果から推定されたNベストのモデル系列に従って、音韻HMMを連結し、その連結した音韻HMMの入力音声に対する尤度が最大となるように、音韻HMMのパラメータを推定して、認識対象話者に適応化する方法が提案されている。この方法での 適応化は、次のステップからなる。

(1) 適応化前の音韻 H M M を用いて入力音声の N ベスト 認識を行い、入力音声に対する N 個のモデル系列を求め る。

(2)各モデル系列ごとに、音韻HMMのパラメータをその話者に適応化する。

(3)適応化後に、最大尤度を示したモデル系列を選択する

(4) その適応化された音韻HMMのパラメータを用いて 現在のHMMを更新する。

30 上記、ステップ(2)~(4)を繰り返す。従って、上記方法は、入力音声の途中でモデルを変更することができない。

【0004】図14は従来のビーム探索を用いる音声認識方式の機能ブロック図である。音声区間切出手段1によって、入力音声1001から切り出された音声区間の各フレームについて、分析手段2による音声分析を行い、特徴パラメータの時系列3を得る。そして、モデル記憶手段51からパラメータ5aを、また構文情報格納手段4から入力音声に対応するモデルの系列を規定する構文情報をそれぞれ参照し、特徴パラメータの時系列3に対する最適なモデル系列を認識結果1003として、以下のようにして得る。なお、10は入力音声の第1フレームから途中までのフレームに対応する部分モデル系列の仮説を格納する部分モデル系列格納手段である。

【0005】最初のフレーム番号を1、最後のフレーム番号をTとする。まず、最初に、部分モデル系列の初期値を部分モデル系列格納手段10に格納する。次に入力音声のフレーム番号iを1とおく。モデル演算手段7は部分モデル系列格納手段10から部分モデル系列の仮説(Hとする)をとり出す。つぎに、構文情報格納手段4

-2-

50

の構文情報から、部分モデル系列Hに連結可能なモデル (音韻モデルk、複数通りのときもある)を選択し、音 韻モデルkに対応するフレーム番号iの特徴パラメータ の尤度f(k,i)を計算する。さらに、音韻モデルkを連結 した1フレーム分成長した部分モデル系列の仮説を作成 し、ビーム探索用の中間スタック1004に格納する。 1フレーム分成長した部分モデル系列の仮説の累積尤度 は成長前の種の部分モデル系列の累積尤度に音韻モデル kの尤度を加えたものである。ビーム探索手段9はフレ ーム番号iについて、中間スタック1004内の部分モ デル系列の仮説の累積尤度を相互に比較し、例えば、累 積尤度の最大の仮説の尤度を上限とし、この上限からビ ーム幅8だけ引いた値を下限として、この範囲の累積尤 度を有する部分モデル系列の仮説を部分モデル系列格納 手段10に格納する。

【0006】この場合、中間スタックからの仮説の選び 方としては、例えば、累積尤度の大きい方からNb個の 部分モデル系列を選ぶこともできる。ただし、Nbはビ ーム内に残す仮説の数の最大の数を表す。以上の処理を 入力音声の第1フレームから最後のフレームまで行うこ とによって、部分モデル系列格納手段10には、入力音 声の全フレームに対応するモデル系列の仮説がその累積 尤度とともに得られる。その後、後向き探索手段100 2は、例えば最適な累積尤度の仮説を選ぶことによっ て、認識結果1003を得る。

#### [0007]

【発明が解決しようとする課題】入力音声の途中でビー ム探索の閾値を変更する従来のビーム探索は、探索量を 削減することができるが、認識に用いるモデルのパラメ ータは一定であり、このようなパラメータが一定のモデ ルで認識を行うため、認識精度の向上は得られない。ま た、従来の教師なし適応化は、一定のパラメータのモデ ルでNベスト認識を行ってN個のモデル系列を求めた後 に、認識結果からモデルのパラメータの入力音声への適 応化を行う。このため、精度のより高い認識結果を得る ためには、適応化された音響モデルにより、再度の認識 処理が必要であるという問題があった。

【0008】この発明が解決しようとする課題は、ビー ム探索を用いる音声認識において、入力音声の途中で音 韻モデルおよび音韻境界のモデルを含むモデルを入力音 40 声に適応化することで、認識の精度を向上させることで ある。また、入力音声の途中でモデルの適応化を行うと ともに、入力音声の途中で得られるモデルの精度改善の 結果として、入力音声の途中でビーム探索の幅を絞るこ とで、探索量を削減することである。

#### [0009]

【課題を解決するための手段】この発明に係る音声認識 方式は、入力音声の各フレームで入力音声にモデルを適 応化し、この適応化されたモデルの系列からなるモデル 系列に対する入力音声の尤度を探索処理により求め、こ 50 る。図3に対応するグラフの構文状態の遷移表は図4の

の尤度に基づき音声認識をするようにしたものである。 【0010】上記探索処理としてはビーム探索が用いら れる。

【0011】また、入力音声の途中のビーム内の仮説の 尤度の大きい方からN個の仮説を用いて、上記モデルを 入力音声へ適応化するようにする。

【0012】また、入力音声の途中のビーム内の仮説の 尤度の大きい方からN個の仮説を用いるとともに、尤度 に応じた重みによる加重を用いて上記モデルを入力音声 へ適応化するようにする。

【0013】また、上記モデルとしては音韻のモデルあ るいは音韻境界のモデルとする。

【0014】また、上記音韻のモデルあるいは音韻境界 のモデルをセミ連続分布のモデルとし、音韻のモデルの 分岐係数だけを適応化する。

【0015】また、上記音韻境界のモデルの入力音声へ の適応化において、入力音声の途中のビーム内の仮説の 尤度の大きい方からN個の仮説を選択する際に、モデル 間の遷移を有する部分モデル系列の仮説を選択するよう にする。

【0016】また、ビーム探索の探索の幅をフレームご とに適応化係数を用いて変化させるようにする。

#### [0017]

20

#### 【発明の実施の形態】

実施の形態1.この実施形態は、モデルとして混合連続 分布の音韻モデルを用いる場合を示す。図1は、この実 施形態における音声認識方式の機能ブロック図である。 入力音声信号1001は音声切出手段1により、例えば 10msの一定の分析周期で、例えば25.6msの信号区 間であるフレームに分割される。分析手段2は、これを フレームごとに特徴パラメータ3に変換する。フレーム 番号 t の特徴パラメータをXtと記す。図2はこれ以降の 動作を示すフローチャートである。ステップ21ではモ デルの初期化を行う。すなわち適応化前のモデルである 初期モデルを初期モデル記憶手段5からモデル記憶手段 6にコピーする。また、フレームの番号 t を 1 に設定す る。次に、t=1番から最終のt=T番のフレームにつ いて、フレーム番号 t を 1 づつ増加しながら、フレーム ごとに以下の処理を行う。

【0018】構文情報格納手段4に格納された構文情報 は、部分モデル系列のあとに接続可能なモデルを決める ための情報であり、状態をあらわすノードと、遷移をあ らわす枝とから表わされる。これは例えば図3に示すよ うなグラフとして表現される。またこの構文情報は、構 文情報格納手段4内においては図4に示すような表とし て格納されている。すなわち、ある部分モデル系列の現 在の構文的な状態をあらわす番号から、次に接続可能な モデルと、そのモデルを選択したときに拡張された部分 モデル系列の次の状態番号が、表として与えられてい

ようになる。

【0019】モデル演算手段7は、1フレーム分の入力を行い(ステップ22)、フレームごとに、構文情報に従ったモデルのパラメータを適用し(ステップ23)、入力される特徴パラメータ3の尤度を計算する(ステップ24)。モデルのパラメータは、音韻モデルkについて、M混合のガウス分布の平均、分散、分岐係数 $\mu$ m(k)、 $\Sigma$ m(k)、 $\lambda$ m(k)(m=1, 2, ..., M) からなる。現在の構文状態が $\mu$ 0とき、構文情報から自己ループを含めて後続の遷移可能なすべての枝を検知し、このすべての枝について、その枝のモデルと遷移先の構文状態の組み合わせ $\mu$ 10に対するモデル $\mu$ 10に対するモデル

[0020]

【数1】

$$f(t,k) = \sum_{m=1}^{M} \lambda_m(k) N(x_t, \mu_m(k), \Sigma_m(k))$$

【0021】ステップ25では、次のようにして、17レーム前の部分モデル系列を17レーム分拡張し、新しい部分モデル系列を生成する。種となる一フレーム前の部分モデル系列が S1, S2, ... のとき、部分モデル系列を一つ選択し、Sとする。Sは構文状態 $\delta$ (S)と、累積 尤度 $\alpha$ (S)と、最終モデルk(S)とを情報として保持している。Sの構文状態がpのとき、つぎの演算を行い、構文状態、選択されるモデルの組み合わせに応じて、新しい部分モデル系列の仮説U1, U2, ... を生成する。例えば、選択されるモデルがkで、次の構文状態がqのとき、これに対応して生成される新しい部分モデル系列をUとすると、Uの構文状態 $\delta$ (U) は $\delta$ (U) = q、O0 以の累積 尤度 $\alpha$ (U) は $\alpha$ (U) =  $\alpha$ (S) + f(U, V) の最終モデルはV0 以の目をである。

【0022】ビーム探索手段9は、モデル演算手段7で生成された部分モデル系列U1,U2,...について、それらの累積尤度と、制御手段1000より与えられるビーム幅8とで決まる、ビーム幅範囲の中に入らない仮説を破棄することで、ビーム幅の中に入る仮説だけを残し、舶エモデル系列として出力し、部分モデル系列格納手段10に格納する(ステップ26)。ビーム幅8に基づくビーム幅範囲の設定は、U1,U2,...の累積尤度の中の最大値 $\alpha$ maxを上限として、 $\alpha$ maxからビーム幅8を減じた値を下限 $\alpha$ minとすることで行う。枝刈りは、U1,U2,...の中から、その累積尤度 $\alpha$ (U1), $\alpha$ (U2),...が、 $\alpha$ minから $\alpha$ maxの間にある仮説を残し、それ以外を破棄することで行う。

【0023】次に部分モデル系列選択手段11は、適応 化手段13における適応化に用いる情報としての部分モ デル系列を選択する(ステップ27)。例えば、部分モ デル系列格納手段10の中の部分モデル系列で、累積尤 度の大きい部分モデル系列から順番に探索し、異なるモ デルを選択した部分モデル系列を最大でN個選択する。

【0024】次に適応化手段13は、部分モデル系列選択手段11が現在のフレームで選択した部分モデル系列U1,U2,... (最大でN個)の、選択されたモデル $k \in k$  (U1),k (U2),... (最大でN個)について、適応化係数12に従って、パラメータの適応化を行う(ステップ28)。この実施形態においては、モデルのパラメータは、音韻モデルkについて、M混合のガウス分布の平均、分散、分岐係数 $\mu$ m(k), $\Sigma$ m(k), $\lambda$ m(k) (m=1,2,...,M)からなる。適応化の対象は、M混合の各分布

(正規密度関数)の尤度に対する分岐係数  $\lambda$  m(k)と、M 30 混合の各分布の平均  $\mu$  m(k)であり、従って、補正前のパラメータ 1 4 は、モデルkについて  $\lambda$  m(k)と、 $\mu$  m(k)であり、その適応化は、次式で行う。

[0025]

【数2】

$$\mu_m(k) \leftarrow \mu_m(k) * (1-w) + \frac{x_t \gamma_m}{M} * w \\ \sum_{m=1}^{\infty} \gamma_m \\ (m = 1, 2, \dots, M)$$

$$\lambda_m(k) \leftarrow \lambda_m(k) * (1-w) + \gamma_m * w$$

$$(m = 1, 2, \dots, M)$$

ここで、

$$\gamma_m = rac{\lambda_m(k)N(x_t,\mu_m(k),\Sigma_m(k))}{\sum\limits_{m=1}^{M}\lambda_m(k)N(x_t,\mu_m(k),\Sigma_m(k))}$$

【0026】なお、wは適応化係数12で0≦w<1。 分散の適応化は理論上は次式で可能であるが、適応化の 対象となるパラメータ数を削減するため、この実施形態\* \*では行わない。 [0027] 【数3】

$$\Sigma_{m}(k) \leftarrow \Sigma_{m}(k) * (1 - w) + \frac{(x_{t} - \mu_{m}(k))(x_{t} - \mu_{m}(k))^{t} \gamma_{m}}{\sum\limits_{m=1}^{M} \gamma_{m}} * w$$

 $(m=1,2,\ldots,M)$ 

【0028】全ての選択されたモデルについて、上記の 適応化が終了した後、適応化手段13は適応化の結果得 られたパラメータを補正後パラメータ15としてモデル 記憶手段6のパラメータを補正後のパラメータ15に置 き換える(ステップ29、30)。そして、後向き探索 手段1002による後向き探索を行い、認識結果100 3を得る(ステップ31)。なお、制御部1000はモ デル記憶手段6の初期化から、入力のフレームに同期し たモデル演算手段7の処理、ビーム探索手段9、適応化 手段13の各処理の制御を行う。

【0029】以上のように、t番目のフレームでの入力 20 フレームの尤度計算に用いるモデルのパラメータは、一 つ前のフレームで適応化処理により補正されたパラメー タを用いている。これにより、次第に適応化が進んでい く。すなわち、認識結果が出たあとではなく、認識処理 中に適応化が進められるものである。また、構文情報を 備えるビーム探索の過程の中で、構文情報で規定される 部分モデル系列から、尤度の高い部分モデル系列のモデ ルを適応化の対象のモデルとして選択しているため、い わば過去の履歴で補正されたフレームごとの認識結果に よるモデルの適応化が実現されることになっている。こ のため、従来のビーム探索のビーム幅の制御による、探 索量の減少の効果とともに、従来は得られなかった認識 精度の改善の効果が期待できる。また、部分モデル系列 選択手段11において、累積尤度の大きい部分モデル系 列から順番に探索し、異なるモデルを選択した部分モデ ル系列を最大でN個選択するようにしたので、安定した 適応化が行える。

【0030】実施形態2.次に、モデルとしてセミ連続 分布の音韻モデルを用いる実施形態を示す。この場合の ブロック図は図1と同じであり、フローチャートは図2 と同じである。モデルが異なるため、モデル演算と適応 化部の動作が異なるが、それ以外は同じであり、説明を 省略する。

【0031】モデル演算手段7は、フレームごとに、構 文情報4に従ったモデルのパラメータを適用し、入力の 特徴パラメータ3の尤度を計算する。この実施形態のモ デルのパラメータは、すべての音韻について共通のM個 10 のコードブックのガウス分布の平均、分散 μ m, Σ m (m= 1, 2, ..., M) と、音韻モデル k についての分岐係数 λ m(k) からなる。現在の構文状態が p のとき、構文情報から自 己ループを含めて後続の遷移可能なすべての枝を検知 し、このすべての枝について、その枝のモデルと遷移先 の構文状態の組み合わせ < k, q > ∈ { < k1, q1 >, < k2, q 2>、...、<kn, qn> ) に対するモデルkの特徴パラメ ータ x tの尤度 f (t, k)を、混合分布の各分布の尤度N(x  $t, \mu m, \Sigma m$ ) の加重和として次式で計算する。

[0032]

【数4】

30

40

$$f(t,k) = \sum\limits_{m=1}^{M} \lambda_m(k) N(x_t, \mu_m, \Sigma_m)$$

【0033】種となる一フレーム前の部分モデル系列が S1, S2,... のとき、部分モデル系列を一つ選択し、S とする。Sは構文状態 $\delta$ (S)と、累積尤度 $\alpha$ (S)と、最終 モデルk(S)とを情報として保持している。Sの構文状 態がpのとき、つぎの演算を行い、構文状態、選択され るモデルの組み合わせに応じて、新しい部分モデル系列 の仮説U1, U2,...を生成する。例えば、選択されるモ デルがkで、次の構文状態がqのとき、これに対応して 生成される新しい部分モデル系列をUとすると、Uの構 文状態  $\delta$  (U) は  $\delta$  (U) = q 、 U の累積尤度  $\alpha$  (U) は  $\alpha$  (U) =  $\alpha$ (S) + f(t, k)、Uの最終モデルはk(U) = kである。

【0034】適応化手段13は、部分モデル系列選択手 段11が現在のフレームで選択した部分モデル系列U1.  $U2, \ldots$  (最大でN個) の、選択されたモデル $k \in k$  (U 1), k (U2), ... (最大でN個) について、適応化係数1 2に従ってパラメータの適応化を行う。この実施形態の モデルkのパラメータは、すべての音韻について共通の M個のコードブック(正規密度関数、平均、分散 Σm, λ m (m=1, 2, ..., M)) である。適応化対象は音韻モデルk についての分岐係数μm(k)である。従って、補正前のパ ラメータ14は、モデルkについてλm(k)であり、その 適応化は次式で行う。

[0035]

【数5】

$$\lambda_m(k) \leftarrow \lambda_m(k) * (1-w) + \gamma_m * w$$

$$(m = 1, 2, \dots, M)$$

$$\gamma_m = rac{\lambda_m N(x_t, \mu_m, \Sigma_m)}{\sum\limits_{m=1}^{M} \lambda_m N(x_t, \mu_m, \Sigma_m)}$$

【0036】なお、 $N(xt, \mu m, \Sigma m)$ が第m番目のコードブックの尤度(正規密度関数の値)である。 $\lambda m=0$ なる分岐係数は、適応化しても $\lambda m=0$ のままである。この実施形態では、したがって、 $\lambda m=0$ なる係数についての適応化のための演算を省略することで、精度に影響を与えずに、演算量を削減することができる。すべてのモデルについて、上記の適応化が終了した後、適応化手段 13は、適応化の結果得られたパラメータを補正後パラメータ 15としてモデル記憶手段 6 のパラメータを補正後のパラメータ 15 に置き換える。

【0037】以上のように、実施形態1と同様、t番目 のフレームでの入力フレームの尤度計算に用いるモデル 20 のパラメータは、一つ前のフレームで適応化処理により 補正されたパラメータを用いている。これにより、次第 に適応化が進んでいく。すなわち、認識結果がでたあと ではなく、認識処理中に適応化が進められるものであ る。また、構文情報を備えるビーム探索の過程の中で、 構文情報で規定される部分モデル系列から、尤度の高い 部分モデル系列のモデルを、適応化の対象のモデルとし て選択しているため、いわば過去の履歴で補正されたフ レームごとの認識結果によるモデルの適応化が実現され ることになっている。このため、従来のビーム探索のビ ーム幅の制御による、探索量の減少の効果とともに、従 来は得られなかった認識精度の改善の効果が期待でき る。この実施形態では、セミ連続分布を用いたため、分 岐係数の適応化だけで精度が改善される。計算、適応化 が容易である。

【0038】実施形態3.次に、音韻のモデルについて、フレームごとに適応化を行うもので、モデル系列の 尤度に応じた適応化係数による適応化をする実施形態を 示す。 \*【0039】この場合のブロック図は図1と同じであ

10 り、フローチャートは図2と同じである。音韻のモデルは、実施形態2と同様のセミ連続分布モデルである。この実施形態では音韻モデルとしてセミ連続分布モデルについて説明したが、混合連続分布モデルでも、同様な効果が期待できる。適応化手段13の動作が異なる以外は実施形態2と同様であり、説明を省略する。

【0040】適応化手段13は、部分モデル系列選択手段11が現在のフレームで選択した部分モデル系列U1、U2、... (最大でN個)の、選択されたモデル $k \in k$  (U1)、k (U2)、... (最大でN個)について、適応化係数12に従って、選択された部分系列の尤度に応じて、パラメータの適応化を行う。モデルk について、適応化係数w (k)の適応化を行う。ここで、モデルkの適応化係数w (k) は、

[0041]

【数6】

$$w(k) = w \times \frac{\alpha(U(k))}{\sum\limits_{k \in \{n \in \mathcal{T}_k\}} \alpha(U(k))}$$

【0042】式中、U(k) は選択されたモデルk を選択するにあたって用いられた部分モデル系列である。この実施形態のモデルkのパラメータは、すべての音韻について共通のM個のコードブック(正規密度関数、平均、分散  $\mu$  m、 $\Sigma$  m (m=1,2,...,M))である。適応化対象は、音韻モデルk についての分岐係数  $\lambda$  m(k) である。従って、補正前のパラメータ 14は、モデルk について  $\lambda$  m (k) であり、その適応化は、次式で行う。

[0043]

【数7】

$$\lambda_m(k) \leftarrow \lambda_m(k) * (1 - w(k)) + \gamma_m * w(k)$$

$$(m = 1, 2, \dots, M)$$

ここで、

$$\gamma_m = rac{\lambda_m N(x_t, \mu_m, \Sigma_m)}{\sum\limits_{m=1}^{M} \lambda_m N(x_t, \mu_m, \Sigma_m)}$$

【0044】なお、 $N(xt, \mu m, \Sigma m)$ が第m番目のコードブックの尤度(正規密度関数の値)である。  $\lambda m = 0$ なる分岐係数は、適応化しても  $\lambda m = 0$ のままである。

この実施形態では、したがって、  $\lambda$  m = 0 なる係数についての適応化のための演算を省略することで、精度に影 50 響を与えずに、演算量を削減することができる。すべて

のモデルについて、上記の適応化が終了した後、適応化 手段13は、適応化の結果得られたパラメータを補正後 パラメータ15としてモデル記憶手段6のパラメータを 補正後のパラメータ15に置き換える。

【0045】以上のように、実施形態1と同様、 t 番目 のフレームでの入力フレームの尤度計算に用いるモデル のパラメータは、一つ前のフレームで適応化処理により 補正されたパラメータを用いている。これは、構文情報 を備えるビーム探索の過程の中で、構文情報で規定され る部分モデル系列から、尤度の高い部分モデル系列のモ デルを、適応化の対象のモデルとして選択しているた め、いわば過去の履歴で補正されたフレームごとの認識 結果によるモデルの適応化が実現されることになってい る。このため、従来のビーム探索のビーム幅の制御によ る、探索量の減少の効果とともに、従来は得られなかっ た認識精度の改善の効果が期待できる。この実施形態で はセミ連続分布を用いたため、分岐係数の適応化だけで\* \*精度が改善される。計算、適応化が容易である。また、 部分系列の尤度を考慮するため、誤った方向の適応化を 防止することが期待できる。

【0046】実施形態4.次に音韻境界のモデルについ て、フレームごとに適応化を行うものを示す。音韻境界 のモデルは、音韻間の遷移に対応したモデル間の遷移を 制御するためのモデルであり、次の尤度比が1より大き いときに音韻間の遷移が可能である。

<尤度比>=<音韻境界である第1の確率密度>/<音 韻境界でない第2の確率密度>

この実施形態では、第1の確率密度(Pr(Bt | 境界))およ び第2の確率密度 (Pr (Bt | 非境界))は、コードブックの 確率度密度関数の次の多項式で与えられる。但し、Bt はt番目及びその前後のフレームから作成した特徴量で ある。

[0047]

【数8】

$$Pr(B_t|$$
境界 $) = \sum\limits_{m=1}^{MB} P_m N(B_t, \mu_m, \Sigma_m)$ 

$$Pr(B_t|$$
非境界 $) = \sum\limits_{m=1}^{MB} Q_m N(B_t, \mu_m, \Sigma_m)$ 

ただし、

$$\sum_{m=1}^{MB} P_m = 1$$

$$\sum_{m=1}^{MB} Q_m = 1$$

$$0 \leq P_m \leq 1(m=1,2,\ldots,MB)$$

$$0 \leq Q_m \leq 1(m=1,2,\ldots,MB)$$

【0048】この実施形態での部分モデル系列選択手段 11は、部分モデル系列格納手段10の部分モデル系列 の中から、音韻境界の遷移が起こった部分モデル系列 (即ち、自己ループに対応しないもの)を尤度の大きい 方から、最大でN個選択する。これにより、特別な計算 をすることなく選択が行える。また、この実施形態での 適応化手段13は、部分モデル系列選択手段11が現在 のフレームで選択した部分モデル系列U1, U2,... (最 大でN個)の、選択されたモデル $k \in k(U1)$ , k(U2),

... (最大でN個) について、適応化係数12に従っ て、パラメータの適応化を行う。この実施形態の音韻境 界モデルkのパラメータは、コードブックの尤度に対す る分子多項式係数 Pm(k) であり、従って、補正前のパラ メータ14はモデル k について Pm(k) であり、その適応 40 化は次式で行う。

[0049]

【数9】

$$P_m(k) \leftarrow P_m(k) * (1-w) + \gamma_m * w$$

$$(m = 1, 2, \dots, MB)$$

ただし、

$$\gamma_m = rac{P_m N(B_t, \mu_m, \Sigma_m)}{\sum\limits_{m=1}^{MB} P_m N(B_t, \mu_m, \Sigma_m)}$$

【0050】なお、MBは音韻境界モデル用のコードブ 10 に基づいて作成した実験結果を示す。 ック (正規密度関数) の数、N(B t, μm, Σm)は正規密 度関数、μm, Σmはそれぞれ正規密度関数の平均および 分散である。wは適応化係数である。 Pm= Oなる多項 式係数は、適応化してもPm=0のままである。この実 施形態では、したがって、Pm=0なる係数についての 適応化のための演算を省略することで、精度に影響を与 えずに演算量を削減することができる。すべてのモデル について、上記の適応化が終了した後、適応化手段13 は、適応化の結果得られたパラメータを補正後パラメー タ15としてモデル記憶手段6のパラメータを補正後の 20 パラメータ15に置き換える。

【0051】実施形態5.次に、フレームごとのモデル の適応化処理とともに、ビーム探索の幅を、フレームに 同期して、斬減させる例を示す。図5にビーム探索の幅 の変化を模式的に示す。フレームごとのモデルの適応化 処理によって、尤度が高くなることが期待され、ビーム 内における正解の仮説の順位が向上する。このため、ビ ーム幅をフレームごとに斬減させることで、探索量が削 減される。ビーム幅8の更新は次式で行う。但し、 $\theta$ は ビーム幅である。

#### θ ← θ \* (1-w) + < ビーム幅推定値 > \* w

【0052】ここで、<ビーム幅推定値>は、数多くの 例について認識実験を行い、最終入力フレームにおいて 正解の部分モデル系列の尤度と、そのときの尤度が最大 の部分モデル系列の尤度との差として求めた。ビーム幅 の初期値は、<ビーム幅推定値>に比べ、大きな値に設 定する。上の式でwはビーム幅をフレームごとに更新す るときの度合いを決める適応化係数である。

【0053】適応化係数をどのように設定するのが妥当 かを実験的に決めるため評価実験を行った。音節の3連 40 鎖の制約を構文情報とする。出力の仮説はグラフ構造に なっている。グラフ構造の複雑さの減少の程度でフレー ムごとの適応化の効果を調べた。図6は、音韻モデル (セミ連続分布モデル) の1フレームごとの適応化を行 う実施形態 2 の適応化のため選択する仮説数 N と適応化 係数wとの組み合わせ条件について、

(1) 正解のモデル系列の尤度と最大の尤度を示したモデ ル系列の尤度との差(Δ)

- (2) 出力グラフのノード数
- (3)出力グラフのエッジ数 (枝の数)

【0054】それぞれの数値は、(1)については、フレ ームごとの適応化なしの場合を0として、それに対する △の増加分を、また、(2)と(3)については、フレームご との適応化なしの場合を1として、それに対する比を、 様々な不特定話者の入力音声を認識したときについて平 均した数値を示す。なお、評価に用いた入力音声は、次 の20フレーズである。

【0055】 (話者): (フレーズの音韻記述)

ec10009 : kaizjoowa dociradesuka

ec10009 : kikaisiNkookaikaNnara tookjootawaano mae desu

ecl0009 : tookjootawaano maedesuka

etl1003 : tookjootawaano maedesu

etl1003 : tookjootawaano mae fuj0003 : koNdono hujujasumini

fuj0003 : cukubani cuite osiete kudasai

fuj0003 : cukubawa

fuj0003 : zjeeaarude kuru baaiwa

kdd1005 : koNdo

30 kdd1005 : oNseekeNkjuukaiga aruNde soreo kikini ik itaiNdesukeredo

mac0003 : kikaisiNkookaikaNdesu

mat1003 : koNdo oNseekeNkjuukaiga aruNde

mat1003 : took jootawaadesu

mit0003 : kanazawano rjokooaNnaisjodesjooka

mit0003 : sinainiwa cjuuooni keNrokueNga arimasu

nec1011 : dokoka mite mitai tokorowa arimasuka

nec1011 : kaNkoopuraNzukurio otasukesimasu

nec1011 : rakuhokuhoomeNto

【0056】また図6の結果をグラフにして表現したも のを図7、図8、図9に示す。それぞれX軸を適応化係 数w、Y軸を適応化のため選択する仮説数Nとしたもの であり、Z軸として図7は上記(1)の $\Delta$ 、図8は上記(2) のノード数、図9は上記(3)のエッジ数をとったもので ある。なお、XY平面上にはZ軸の等高線を示してい る。図6~図9から、w=0.005かつN=1~50、また、 w = 0.01かつ $N = 1 \sim 50$ 、さらに、w = 0.02かつ $N = 1 \sim$ 200、w=0.05かつN=50~100にすれば、Δが減少かつ ノード数とエッジ数が減少することがわかる。 Δの減少 50 は音声認識の精度の向上を示し、またノード数とエッジ

数の減少は、音声認識の精度の向上によって、正解以外 のモデル系列の生成が抑制されたことを示していると考 えられる。

【0057】図10は、音韻境界のモデル(セミ連続分 布モデル)の1フレームごとの適応化を行う実施形態4 の適応化のため選択する仮説数Nと適応化係数wとの組 み合わせ条件について、

(1) 正解のモデル系列の尤度と最大の尤度を示したモデ ル系列の尤度との差(△)

(2) 出力グラフのノード数

(3) 出力グラフのエッジ数(枝の数)

に基づいて作成した実験結果を示す。

【0058】それぞれの数値は、(1)については、フレ ームごとの適応化なしの場合を0として、それに対する Δの増加分を、また、(2)と(3)については、フレームご との適応化なしの場合を1として、それに対する比を、 様々な不特定話者の入力音声を認識したときについて平 均した数値を示す。評価に用いた入力音声は、上記の2 0フレーズである。また図10の結果をグラフにして表 現したものを図11、図12、図13に示す。それぞれ 20 X軸を適応化係数w、Y軸を適応化のため選択する仮説 数Nとしたものであり、Z軸として図11は上記(1)の Δ、図12は上記(2)のノード数、図13は上記(3)のエ ッジ数をとったものである。なお、XY平面上には2軸 の等高線を示している。

【0059】図10~図13から、音韻境界モデルの適 応化係数wと適応化する境界の種類数Nの適切な範囲と しては、w=0.1かつ $N=100\sim500$ 、また、w=0.2かつ N = 100、さらに、w = 0.3かつ $N = 50 \sim 500$ 、w = 0.4か  $ON = 50 \sim 500$ 、w = 0.5か $ON = 1 \sim 500$ などで、 $\Delta$ が減 30 の変化を示す説明図である。 少かつノード数とエッジ数が減少することがわかる。Δ の減少は、音声認識の精度の向上を示し、また、ノード 数とエッジ数の減少は、音声認識の精度の向上によっ て、正解以外のモデル系列の生成が抑制されたことを示 していると考えられる。

#### [0060]

【発明の効果】以上に説明したように、この発明によれ ば、入力音声の各フレームで入力音声にモデルを適応化 し、この適応化されたモデルの系列からなるモデル系列 に対する入力音声の尤度を探索処理により求めるように 40 したので、過去の履歴で補正されたフレームごとの認識 結果によるモデルの適応化が実現されることになり、探 索量の減少の効果とともに認識精度の改善の効果があ る。

【0061】また、入力音声の途中のビーム内の仮説の 尤度の大きい方からN個の仮説を用いて、上記モデルを 入力音声へ適応化することにより、安定した適応化が行

【0062】また、入力音声の途中のビーム内の仮説の 尤度の大きい方からN個の仮説を用いるとともに、尤度 50 1 音声区間切出手段 16

に応じた重みによる加重を用いて上記モデルを入力音声 へ適応化することにより、安定した適応化が行える。

【0063】また、上記モデルを音韻のモデルのセミ連 続分布のモデルとし、音韻のモデルの分岐係数だけを適 応化することにより、計算処理や適応化を容易に行うこ とができる。

【0064】また、上記モデルを音韻境界のモデルのセ ミ連続分布のモデルとし、音韻境界のモデルの分岐係数 だけを適応化することにより、計算処理や適応化を容易 10 に行うことができる。

【0065】また、音韻境界のモデルの入力音声への適 応化において、入力音声の途中のビーム内の仮説の尤度 の大きい方からN個の仮説を選択する際に、モデル間の 遷移を有する部分モデル系列の仮説を選択することによ り、容易に選択を行える。

【0066】また、ビーム探索の探索の幅をフレームご とに適応化係数を用いて変化させることにより、ビーム の幅をモデルの適応の度合いに応じて絞り、探索量を削 減することができる。

【図面の簡単な説明】

この発明の実施形態における音声認識方式の 【図1】 機能ブロック図である。

【図2】 この発明の実施形態における音声認識動作の フローチャートである。

【図3】 この発明の実施形態における構文制御情報の 模式図である。

【図4】 この発明の実施形態における構文制御情報の 構成の説明図である。

この発明の実施形態におけるビーム探索の幅 【図5】

【図6】 この発明の実施形態における評価結果の説明 図である。

【図7】 この発明の実施形態における評価結果をグラ フ化して示す説明図である。

【図8】 この発明の実施形態における評価結果をグラ フ化して示す説明図である。

【図9】 この発明の実施形態における評価結果をグラ フ化して示す説明図である。

【図10】 この発明の実施形態における評価結果の説 明図である。

【図11】 この発明の実施形態における評価結果をグ ラフ化して示す説明図である。

【図12】 この発明の実施形態における評価結果をグ ラフ化して示す説明図である。

【図13】 この発明の実施形態における評価結果をグ ラフ化して示す説明図である。

【図14】 従来の音声認識方式の機能ブロック図であ

【符号の説明】

17

- 2 分析手段
- 3 特徴パラメータ
- 4 構文情報格納手段
- 5 初期モデル記憶手段
- 6 モデル記憶手段
- 7 モデル演算手段
- 8 ビーム幅
- 9 ビーム探索手段
- 10 部分モデル系列格納手段
- 11 部分モデル系列選択手段

12 適応化係数

13 適応化手段

14 補正前パラメータ

15 補正後パラメータ

1000 制御手段

1001 入力音声

1002 後向き探索手段

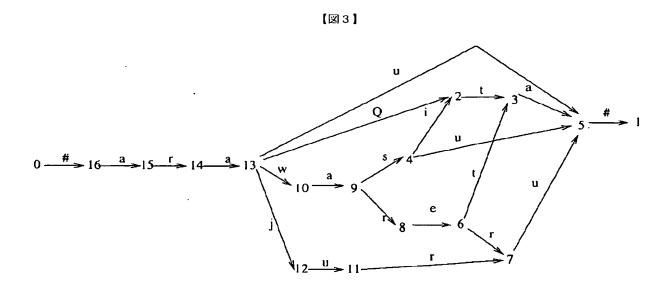
1003 認識結果

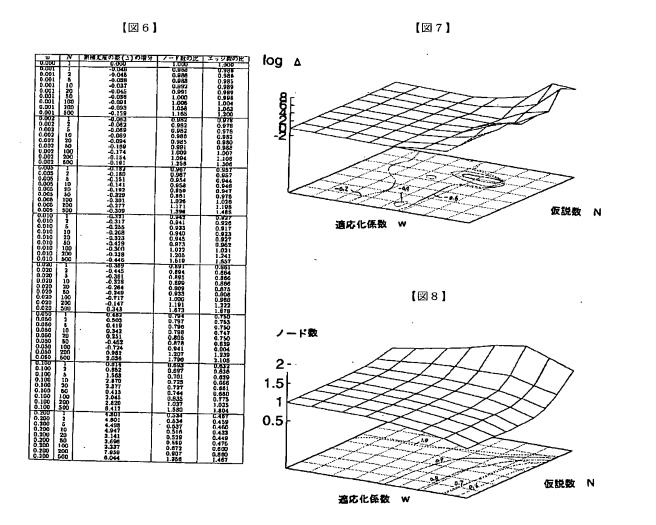
1004 中間スタック

10

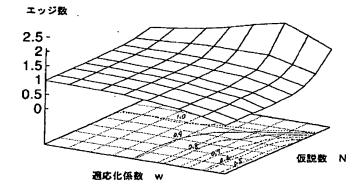
【図2】 【図4】 初期状態の番号 0 最終状態の番号 1 モデル初期化 -21 現在の状態番号 環状される音韻モデル 次の状態番号 *-*22 3 1フレーム入力 ·23 モデル選択 モデル演算 10 5 11 尤度計算 24 7 12 11 13 12 部分モデル系列生成 13 10 13 5 13 14 13 15 14 ビーム探索 16 15 16 -27 部分モデル系列選択 -28 【図5】 適応化 29 ビーム幅の初期値 モデルパラメータ補正 最終ビーム幅 ピーム幅 -30 いいえ Tフレーム終了? はい 0 後向き探索 フレーム番号

[図1]





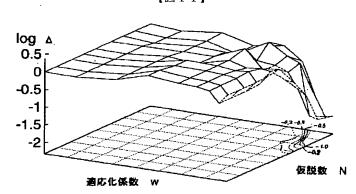
【図9】

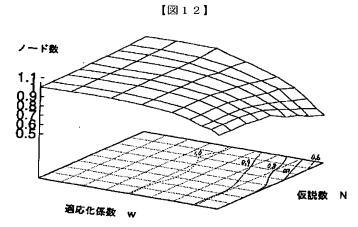


【図10】

	- א	表現之間の変(2)の複分	ノード表の比	エッフ数の比
0.000		0.000	1.000	1,000
0.020	;	-0.013	1.008 1.008 1.010	1.004 1.004 1.009
0.020	Š	-0.013	1.010	1.009
	10	-0.013	1.015	1.017
0.020 0.020 0.020	50	-0.013 0.000	1.012 L020	1.012
0.020	28888	-0.063	1 1.033	l 1.036
0.020	200 500	-0.062	1.032 1.051	1.028
0.030	800	-0.062 -0.062 -0.015	1.031	1.058 1.058 1.008
0.050	î		1.011	1.008
1 0.050	.5.	880.0 880.0	1 1014	1.012
0.050	10	0.088	1.030	1.023
0.050 0.050 0.050 0.050 0.050	20 50	0.088 0.028 0.027 0.025	1.014 1.017	1.019
0.050	ı ım	0.027		1.017
0.050	200		1.020 1.044	1.023
0.100		0.329 0.129 0.367 0.367 0.267 0.026 0.026		0.990
0.100	\$	0.329	0.994 0.995 0.998 0.998 0.998	0.989
0.100	10	0.267	0.995	
0.100	20	0.026	0.998	0.989
0.100	20 50	0.026	0.891	0.989
0.100	100		0.965 0.977	0.962 0.970
0.100 0.100 0.100 0.100 0.100 0.100 0.100 0.100 0.100 0.100	100 200 500	-0.202 -0.175	0.958	0.941
0.200		0.249 0.249	II.UEK	0.950 0.950
0.200	3	0.249	0.988 0.987	0.950
0.200 0.200 0.200	10	0.249	0.967	0.954 0.954
0.200	10 20 50	0.249	0.949	0.933 0.912
0.200 0.200 0.200	100	0.37£ -0.112	0.930	0.912
0.200	700	0.254	0.885	0.852
0.200	500	-0.048 U.276	0.856 0.924	0.825
0.300	1 2	0.226	0.921	828.0
0.300	2	0.226 0.226	0.924	0.901 0.898
0.300	10 20	0.181 0.106	0.921	C.898 C.871
10,300	50	-0.147	O.ATO	0.835
0.300	100	-0.147 -0.164	0.844 0.832	0.835 0.804 0.799 6.737
0.300	200 500	-0.052 -0.401	0.837	0.799
0.400		0.167 0.167	0.875	
0.400	1 3	0.167	0.875	
0.400	10	0.167 0.210	0.579	0.849
0.400	10 20	0.137	0.851	0.849 0.819 0.809 0.763
D.400	100	-0.070 -0.304	0.812 0.773	0.763 0.772
0.400 0.400 0.400 0.400 0.400	100 200 500	-0.304 -0.201	0.732	0.692
0.400	500	-1.323	0.682	I 0.640
0.400 0.500 0.500	7	-0.150 -0.150 -0.150	0.859	0.832 0.832 0.833
0.500	5	-0.150	0.860 0.860	0.832
0.500	10	-0.121	0.846	0.812
0.500 0.500 0.500 0.500	20 50	0.106 -0.347	0.525 0.754	0.785 0.709
0.500	100	_A 749	9.707	0.700
0.500	100	-1.656 -1.860	0.666	0.611
0.500	500	-1.860	0.592	0.540
0.600	ايا	0.113	0.810 0.808	0.770 0.767
0.600 0.600	2 5	0.106 0.137	0.810	0.769 0.748
0.600	i ta i	0.137	0.794	0.748
0.600	20 50	0.203 -0.366	0.790 0.699	0.742 0.436
0.600	100	-0.685	0.666 0.599	0.604
0.600	200 500	-1.710 -1.814	0.599 0.540	0.546
(0.000)	ann l	71.819	U.56U	0.481

【図11】





【図13】

